TIEGCM 集合卡尔曼滤波同化模型设计及初步试验

张亚楠[1,2]　吴小成[1]　胡雄[1]

1 （中国科学院国家空间科学中心　北京 100190）

2 （中国科学院大学　北京 100049）

摘　要 本文选择参数化的电离层热层理论模型 TIEGCM 作为背景模型，基于 COSMIC 掩星观测的电子密度廓线数据，应用集合卡尔曼滤波方法建立了全球电离层电子密度同化模型，实现了全球电离层的电子密度同化。同化结果表明，该同化模型能将观测资料有效地同化到背景模式中，获得全球三维电离层电子密度。与背景模式相比，同化得到的电子密度相对于观测值的偏差显著下降。此外，分组同化与同时同化的结果对比显示，平均偏差改善基本一致，同时同化后的标准偏差在峰值高度以上略有减小。

关键词 电离层，数据同化，集合卡尔曼滤波，同时同化

TIEGCM Ensemble Kalman Filter assimilation model design and preliminary results

ZHANG Ya'nan[1, 2]　WU Xiaocheng[1]　HU Xiong[1]

1 （*National Space Science Center, Chinese Academy of Sciences, Beijing* 100190）

2 （*University of Chinese Academy of Sciences, Beijing* 100049）

Abstract Parameterized ionosphere model TIEGCM is used as background model. Basing on the COSMIC observations, global ionospheric electron density assimilation model is established using ensemble Kalman filter. Result shows, this model can effectively assimilate observations into background model and acquire three dimensional ionospheric electron density. Compared to background, the error between analysis and observations decreases significantly. The root mean square error(RMSE) of NmF2 decreases about 60% for observations assimilated, and 20% for observations not assimilated. The RMSE of hmF2 doesn't get improvement except for mean error.

The results of simultaneous assimilation (SA) and batches assimilation (BA) are compared for this case. The time that the two methods spend in assimilation is about 6 to 7 minutes, which does not differ very much. SA needs nearly 8 GB storage while

---

基金项目 国家自然科学基金资助项目（项目编号：41204137）

作者简介 张亚楠，男，在读博士生，主要从事电离层同化预报研究，E-mail: zhangyanan113@mails.ucas.ac.cn

BA less than 2GB. The statistic of electron density error shows that they nearly acquire the same mean error, but the SA gets relative better improvement in RMSE above 250km.

Key words ionosphere, data assimilation, ensemble Kalman filter, simultaneous assimilation

1 引言

电离层的电子密度分布可以影响导航、短波通讯系统等，以致影响人类活动。电离层受内部和外部驱动因素影响，随经度、纬度、高度、地方时、季节、太阳活动及地磁活动的变化而变化[1]。由于观测缺乏足够的时空分辨率，仅通过观测难以监测电离层状态。

电离层模式一直以来是进行电离层研究和预报的基本工具，包括经验模式和理论模式。经验模式基于大量的观测数据进行统计分析得来，可以较准确的进行中长期预报，很难进行准确的短期预报。理论模式基于物理方程进行建模，可以进行短期预报，但是由于理论模式不可能全面准确的考虑电离层里面的物理过程及扰动情况，再加上边界条件和初始条件的不确定性，及数值方法带来的误差，理论模式进行的预报常常与实际情况有很大的差别[2]。大部分经验模式和理论模式可以模拟电离层气候特征，但由于其缺乏对电离层驱动因素的可靠估计，不能准确的进行电离层现报、预报[1]。

为了满足对电离层进行监测和预报的需求，在气象上已经运用成熟的技术手段——数据同化技术，被人们引用到电离层研究和预报领域。在过去的十年里，基于经验或理论模式和多种电离层观测，许多电离层数据同化模型被开发出来。由于这些模型主要同化有限空间覆盖的地基观测，因此并不能精确地表征全球范围内的电子密度分布。但随着 GNSS 的发展，地基 GNSS-TEC 观测和天基掩星观测的空间覆盖越来越广。尤其是掩星观测，可为海洋、沙漠、极区等其他观测较少的区域提供三维电子密度分布。

国际上电离层数据同化研究从上个世纪就已经开始，其中比较著名的，如美国喷气动力实验室—南加州大学全球同化电离层模式（JPL/USC GAIM）和美国犹他州立大学全球电离层观测同化模式（USU GAIM）[3]。USU GAIM 的同化模式又包括 Gauss-Markov Kalman 滤波同化模式和简化状态 Kalman 滤波同化模式[4, 5]。近年来国外一些学者在电离层同化、电离层驱动参数以及电离层热层的相互作用对电离层同化的影响等方面也开展了一些研究[6, 7]，取得了不少成果。国内这方面的研究还不是很多，主要是乐新安博士先后开展了中低纬电离层模拟和数据同化研究，其在电离层电子密度空间相关性方面的工作在电离层同化研究中具有很大的应用价值[2]。

本文将采用集合卡尔曼滤波同化方法，把 COSMIC 电子密度廓线同化到 NCAR 开发的 TIEGCM 模型中，给出初步结果，并作进了一步讨论，同时对两种同化方式进行了比较。

2 同化模型构建方法

2.1 集合卡尔曼滤波方法

Evensen[8]等提出基于蒙特卡洛随机抽样的卡尔曼滤波方法，也就是集合卡尔曼滤波（Ensemble Kalman Filter, EnKF）。相对于变分方法和卡尔曼滤波，EnKF 采用非线性集合预报的有限样本来估算误差协方差，它可以应用在高度非线性的动力模型中，既不需要切线性

算子或伴随方程，也不需要随时间向后积分。其基本滤波公式如下：

$$P_e = \frac{X'X'^T}{N-1} \quad （1）$$

$$K = P_e H^T (H P_e H^T + R_e)^{-1} \quad （2）$$

$$X^a = X^b + K(Y - HX^b) \quad （3）$$

式中X'为背景扰动场；P_e是背景误差协方差矩阵；R_e是观测误差协方差矩阵。K为增益矩阵，用来修正观测点附近的背景场。X^b为背景场；Y为观测向量；H是观测算子；X^a为同化结果；EnKF 使用样本统计计算背景误差协方差，容易产生虚假的伪相关，因此需要用相关距离来控制伪相关，即协方差局地化。

2.2 同化变量的设置

由（3）式可知，同化所需变量有背景场X^b、观测值Y、由背景场映射到观测值的插值算子或观测算子H以及背景场、观测值的误差协方差矩阵P, R。下面介绍这几个量的获取方法。

背景场由 TIEGCM 模式预报得到。NCAR 开发的 TIEGCM[7]是一种三维、时变、地球热层电离层耦合的理论模型。该模型使用有限差分技术，自洽求解三维中性成分和离子的动量方程、能量方程和连续性方程。其默认设置的经度、纬度方向上的分辨率分别是 5 度和 5 度，高度方向有 29 层，间隔为半个标高，下边界约为 97km，上边界随太阳活动不同在 500km 到 700km 之间变化。

TIEGCM 模式是一种电离层热层电动力学大气环流模型，其包含很多大气参量的预报，如电子密度、大气温度等。这里选取 2006 年第 80 天的数据作为初值，通过模式预报到第 159 天 0 时 0 分。然后扰动模式输入参数中的太阳活动性指数 F10.7，构造均值为 150、标准偏差为 50%的 100 个样本集合。这 100 个样本作为输入参数输入到模式中，模式预报 6 个小时，得到背景场X^b。

观测值为 COSMIC 掩星电子密度廓线数据。这里所用的 COSMIC 掩星观测的电子密度廓线数据是 UCAR 提供的二级电离层数据产品。研究表明，由于 Abel 反演中的球对称假设，F 区高度以下的电子密度的反演结果应当谨慎使用[9]。所以本文只对 150km 以上的电子密度进行同化，插值网格间距为 10km。取 2007 年第 159 天 6 点前后半小时的观测数据作为观测场，对观测数据的质量控制，取观测值大于零的观测。经初步质量控制后约有 3580 个观测点，共 121 条廓线。

TIEGCM 输出的网格高度是等位势高度，为了便于构造观测算子，这里将位势高度换算为几何高度，并插值到等几何高度面。观测算子由观测点所在网格立方体空间线性插值得到。

背景场的误差协方差由集合样本统计得到。在大部分同化系统中，观测误差假设是相互独立的[10, 11]，本文中也采取同样的假设，对观测值不考虑协方差，假设观测偏差为观测值的 20%，则观测误差协方差如下所示：

$$R = \begin{cases} 0.04 \times y_i \times y_i, & i = j; \\ 0, & i \neq j \end{cases} \quad （4）$$

2.3 局地化

在集合卡尔曼滤波方法中，有限集合数造成不满秩和估计的相关虚假噪音。为了消除距观测较远的伪相关，需采取协方差局地化。一般假设电离层空间相关性满足高斯分布，但相关距离目前并没有很好的结论[12]。电离层的相关距离可以看作是各向异性的。这里把电离层的相关性简单的分解为子午方向、纬线圈方向和高度方向的乘积[2]。电离层的相关尺度在

不同的方向有显著不同，且有明显的纬度、高度、地方时、季节、太阳活动变化。垂直方向的相关距离随高度升高而增大，且向上的相关距离大于向下的相关距离，总体上来说，纬向相关距离比子午向相关距离大。

把集合样本估计的背景误差协方差和局地相关函数相乘，得到新的协方差矩阵。

$$\rho_{ij} = e^{-\alpha\times\left(\varphi_{\mathrm{ij}}{}^2/L_x{}^2\right)} \times e^{-\alpha\times\left(\theta_{ij}{}^2/L_y{}^2\right)} \quad (5)$$

$$P_{\mathrm{ij}}^{'} = \rho_{ij} \times P_{ij} \quad (6)$$

ρ_{ij}为i,j两点的相关系数，φ_{ij}、θ_{ij}分别表示两个点在经度和纬度方向的实际距离。若定义两个点之间的相关系数为 0.75 时的距离为相关距离，根据乐新安等[2]人的研究，纬度方向的相关距离L_x取 10°，经度方向的相关距离L_y从中纬~40°变化到赤道地区~20°，高度方向设置不相关。α为相关距离系数，$0.75 = e^{-\alpha}$,α为 0.2877。

2.4 同化方式

顺序同化是对观测数据逐个进行同化，同时同化是同时对所有观测数据进行同化。顺序同化是针对观测数据的局部优化，同时同化是针对所有观测数据的全局优化。研究表明[13]，当观测误差不相关时，在集合卡尔曼滤波中顺序同化和同时同化差异不大。顺序同化每次只同化一个观测数据，对计算机的内存大小要求低，但将所有观测数据都同化完，所需时间长。同时同化一次同化所有观测数据，计算效率高，但要求计算机的内存大，被同化的数据量大时，内存大小会无法满足计算需求。将顺序同化和同时同化的优点相结合，对观测数据进行分批同化，即为分组同化。

为比较分组同化与同时同化的差异，本文分别用这两种方法进行同化，并对同化结果进行了对比分析。本文进行分组同化时，采取逐个观测廓线成批同化的方法。

3 同化实验结果

3.1 同化结果

在 121 条观测廓线中随机取出 50%进行同化，同化方式为同时同化，利用剩余 50%的未参与同化的廓线来验证同化效果，下面给出了一些结果。

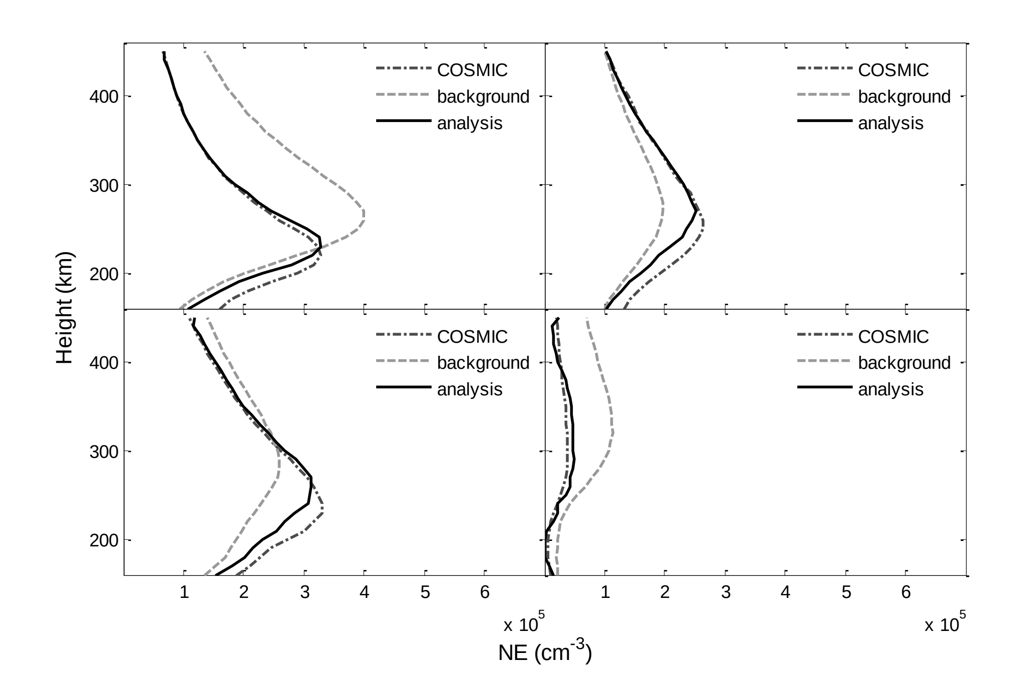


图 1 同化前后背景场、分析场与参与同化的电子密度廓线的比较。点画线为观测场、虚线为背景场、实线为同化结果。

Fig.1Comparison of electron density profile between background (dash lines) and EnKF assimilated (solid lines) for assimilated observations (dash-dotted lines).

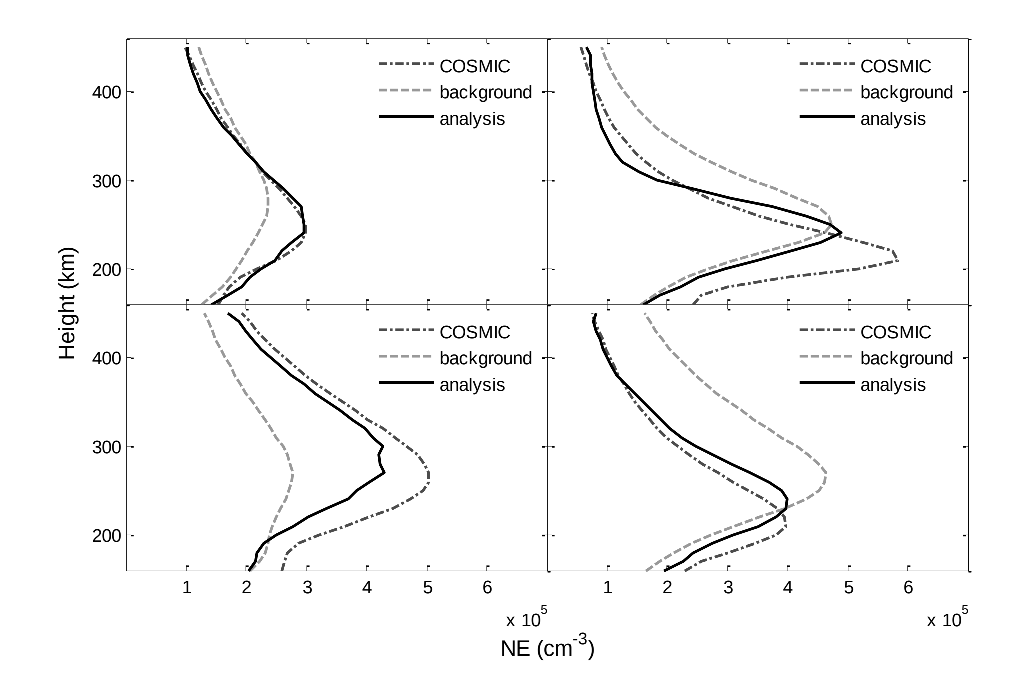


图 2 同化前后背景场、分析场与未参与同化的电子密度廓线的比较。

Fig.2Comparison of electron density profile between background (dash lines) and EnKF assimilated (solid lines) for unassimilated observations (dash-dotted lines).

这里给出了同化结果与观测场及同化之前的背景场的比较个例，深灰色点画线为观测场、浅灰色虚线为背景场、黑色实线为分析场。参与比较的背景场和分析场廓线为插值到相应观测点处的结果。从同化结果可以看出，图 1 中分析场与观测值更为接近，图 2 中虽然这些观测廓线并未参与同化，但分析场仍与观测场符合得更好，这是由背景场提供的相关性所带来的同化效果，说明该同化模型能够将观测资料有效地同化到背景模式中，获得全球三维电子密度分析场，对没有观测资料的地方也能有很好的改善作用。

3.2 相对偏差

为进一步验证同化效果，统计了同化前后在不同高度上背景场和同化结果相对观测的偏差。相对偏差、平均偏差和和标准偏差公式分别为：

$$e_i = \frac{x_i - y_i}{y_i} \times 100\% \tag{7}$$

$$\bar{e} = \frac{\sum_{i=1}^{n} e_i}{n} \tag{8}$$

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}(e_i - \bar{e})^2}{n-1}} \tag{9}$$

其中，x_i为某位置上的背景场或同化结果，y_i为对应位置的观测值，e_i为电子密度的相对偏差，$\bar{e}$和σ_x分别为对应高度上电子密度相对偏差的平均偏差和标准偏差。

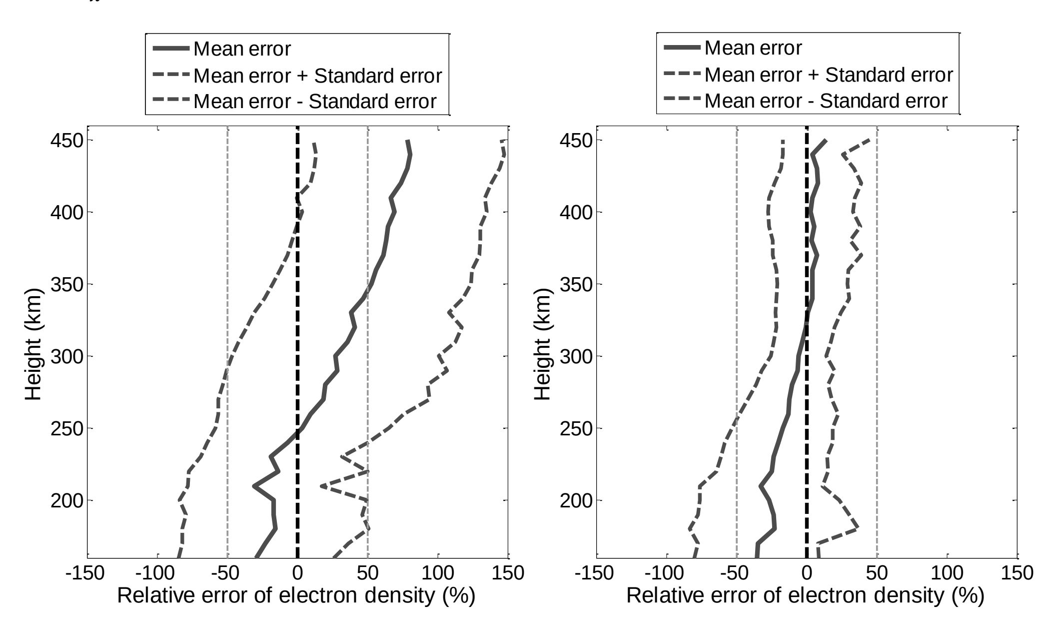


图 3 同时同化前后参与同化观测廓线处的各个高度上电子密度偏差统计（左图为同化前，右图为同化后）

Fig.3.The statistic of electron density error at observation profiles assimilated for simultaneous assimilation. (Left: before assimilation, right: after assimilation).

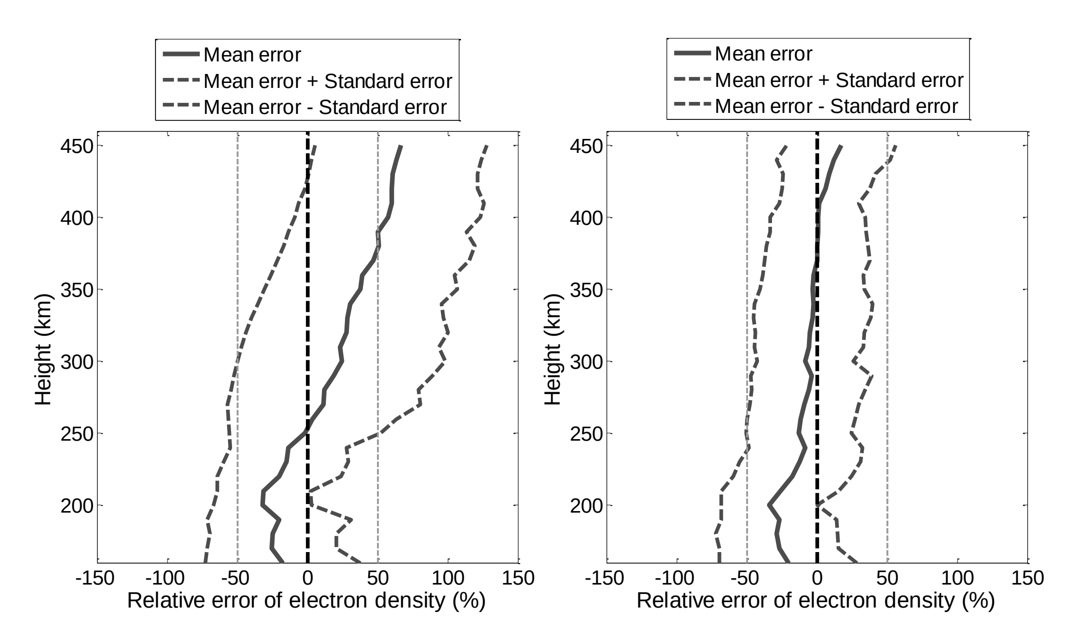


图 4 同时同化前后未参与同化观测廓线处的各个高度上电子密度偏差统计（左图为同化前，右图为同化后）

Fig. 4.The statistic of electron density error at observation profiles unassimilated for simultaneous assimilation. (Left: before assimilation, right: after assimilation).

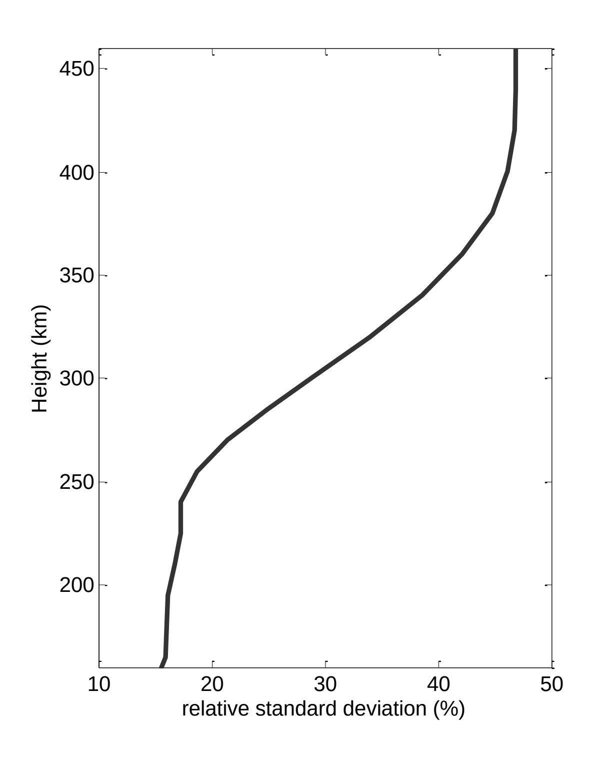


图 5 东经 115 度、北纬 40 度处背景场样本的标准偏差相对于样本均值的百分比

Fig. 5 The profile of the relative standard deviation of background on 115°E 40°N.

图 3、图 4 左图中的实线为同化前背景场的平均偏差，两条虚线分别为平均偏差加减一倍的标准偏差。右图为同化后分析场的偏差统计。可以看出无论是参与同化还是未参与同化的观测廓线处，总体上同化后分析场相比同化之前背景场的平均偏差和标准偏差均有所减小，在 300km 以上平均偏差由同化前的 50%降低到同化后的 15%以内，标准偏差由同化前的 50%

以上降低到同化后的 40%以内。从两图中也发现，在较低高度同化效果并不理想，一是由于此高度范围内的背景值偏小，另一个原因是背景场的样本集合在此范围内样本偏差较小造成的。图 5 给出的是东经 115 度、北纬 40 度处背景误差协方差对角线值的平方根相对于集合样本平均值的百分比，从图中可看出，250km 以下由集合样本统计获得的背景误差在 20%以下，远小于 350km 以上的背景误差。背景值与相对偏差偏小导致背景误差偏小，同化结果就会更接近背景场，与观测数据之间的偏差改善就会大大减小。

3.3 相关性比较

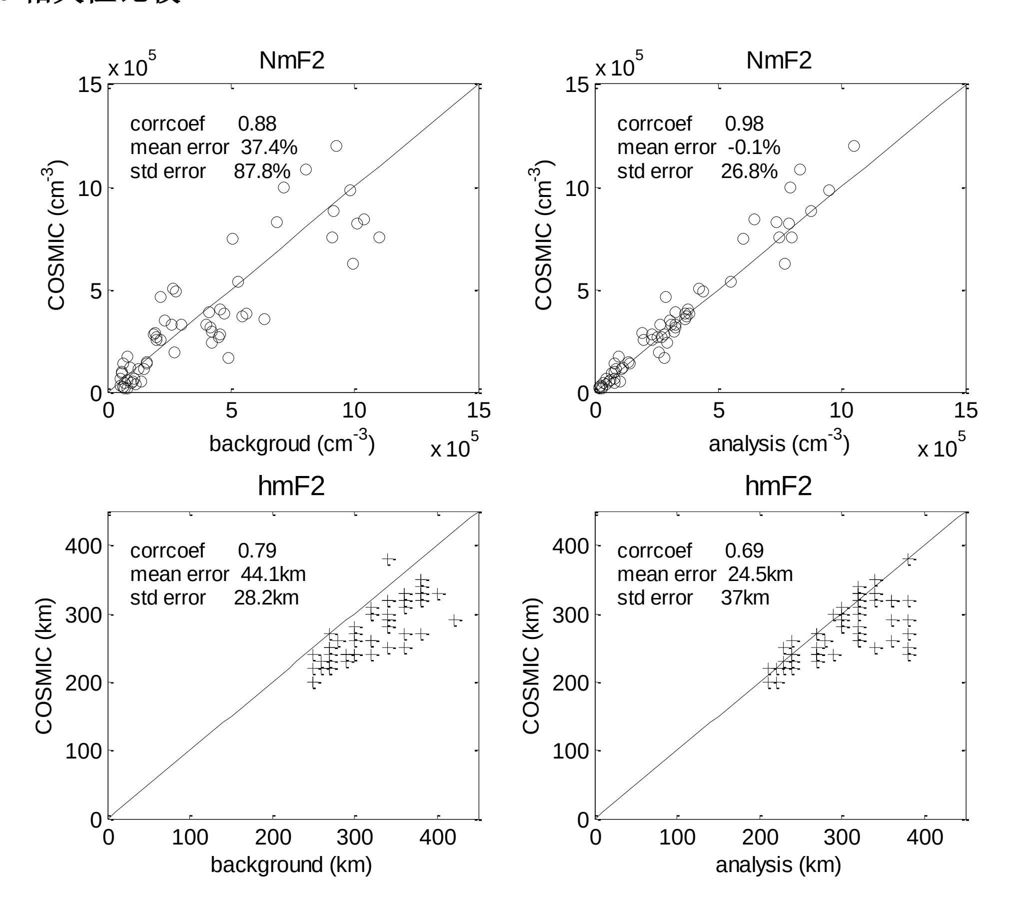


图 6 同化前后参与同化观测廓线处 NmF2 和 hmF2 的相关性统计

Fig.6 Correlation statistic in NmF2 and hmF2 (with assimilation).

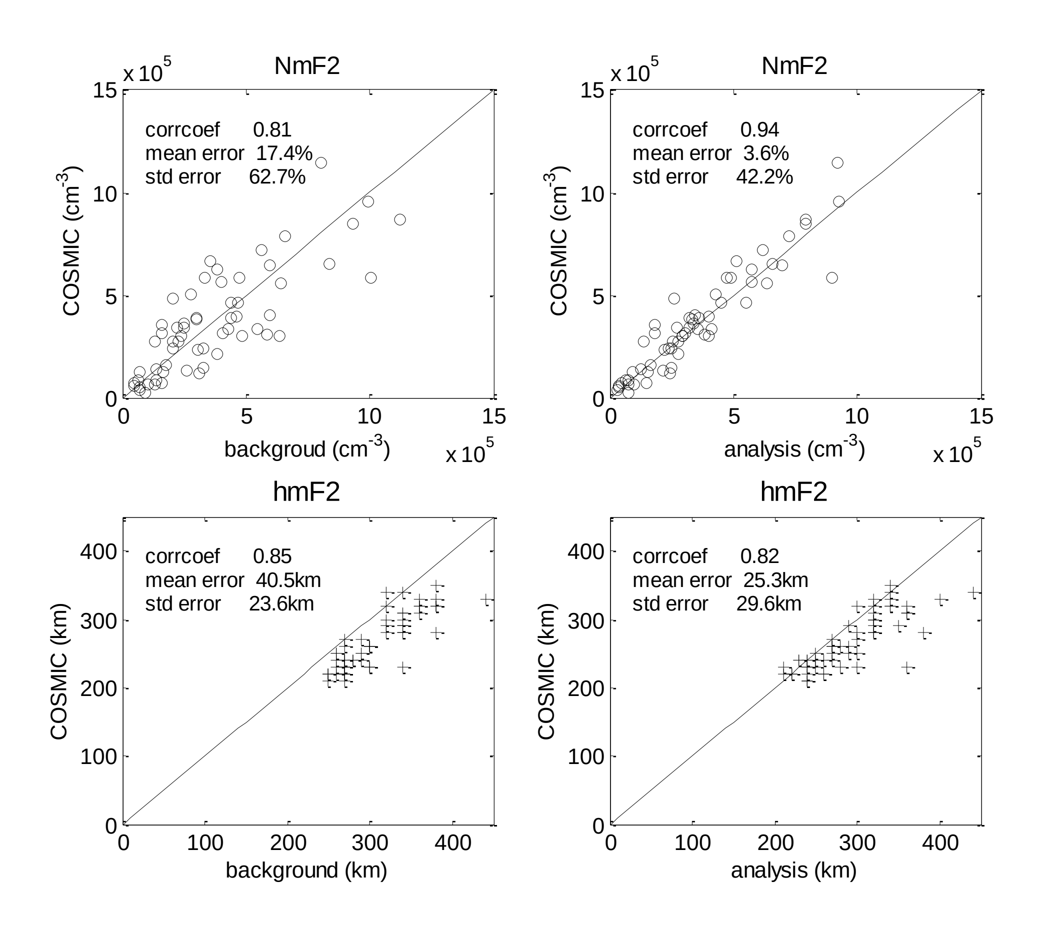


图 7 同化前后未参与同化观测廓线处 NmF2 和 hmF2 的相关性统计

Fig.7Correlation statistic in NmF2 and hmF2 (with no assimilation).

图 6、图 7 中分别给出了电离层峰值电子密度 NmF2 同化前的背景场与观测值、同化后分析场与观测值的散点图。左图横坐标为背景场峰值电子密度，纵坐标为观测值；右图横坐标为分析场，纵坐标为观测数据。从图中可以看出对于参与同化的观测，同化前的背景场与观测值的相关系数为 0.88，同化后为 0.98，同化后相关系数有很大提高，平均偏差从同化前的 37.4%下降到-0.1%，标准偏差降低约 60%。对于未参与同化的观测，相关系数从同化前的 0.81 提高到 0.94，平均偏差从同化前的 17.4%减小到 3.6%，标准偏差减小约 20%。从图中峰值高度比较结果来看，峰值高度同化结果并不理想，平均偏差有所下降，但相关系数和标准偏差并未改善。

3.4 同时同化与分组同化

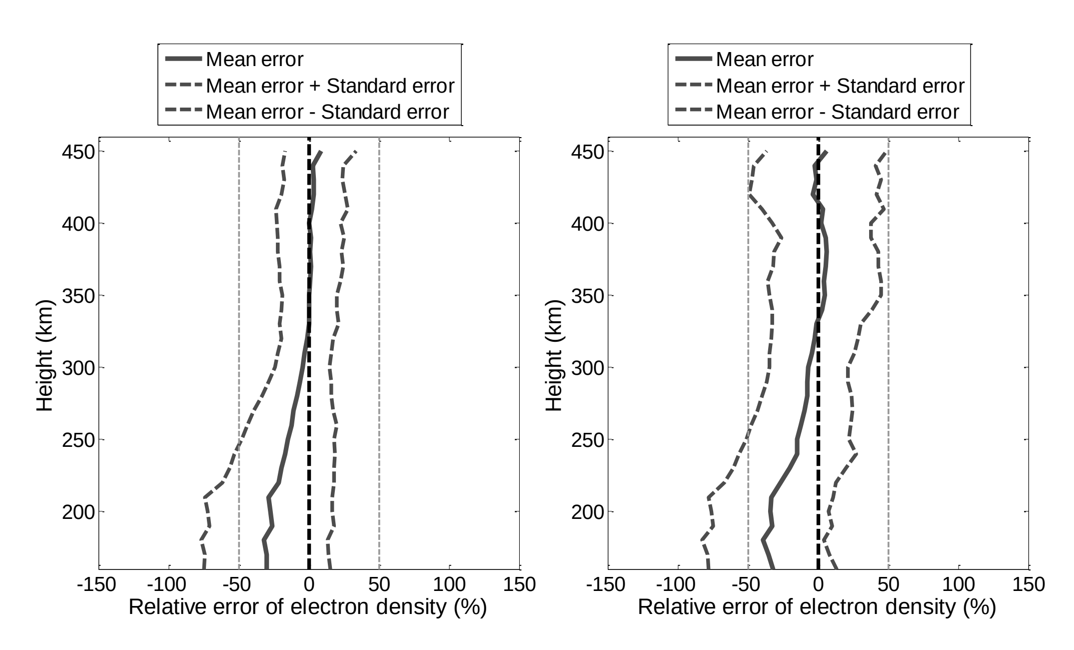


图 8 同时同化后各个高度上电子密度误差统计与分组同化对比。(左图为同时同化，右图为分组同化)

Fig.8 Comparison of the statistic of electron density error between different assimilation methods; the left is simultaneous assimilation and right batches assimilation.

将所有观测数据同时同化到背景场中需要很大的内存存储，该试验是在多核计算机上进行的，CPU 主频为 1.87GHz，同化的观测点数为 3579。同时同化最大占用内存约 8GB，同化耗时 6 分 36 秒。为比较同时同化与分组同化的差异，对同样的观测数据采用分组同化进行了同化实验，而分组同化占用内存 2GB 以内，耗时 7 分 21 秒。如图 8 所示，同化结果对比发现，同时同化和分组同化相对观测值的平均偏差基本一致，但同时同化后的标准偏差在 300km 高度以上相对更小。由此可见，同时同化具有全局优化的优点，但占用内存大；分组同化能够取得不错的同化效果，且占用内存小。

4 结论与展望

本文给出了一个基于集合卡尔曼滤波技术的电离层数据同化模型的初步构建和试验结果。我们利用集合卡尔曼滤波同化方法，将 COSMIC 掩星观测的电子密度廓线同化到 TIEGCM 背景模型中。同化结果表明，采用集合卡尔曼滤波算法，把电子密度同化到背景模式中能够获得较好的同化结果，对于参与和未参与同化的观测，NmF2 的标准偏差分别降低约 60%和 20%，说明我们设计的算法可行、所选择的各种参数比较合理。分组同化与同时同化的对比结果表明，分组同化对内存需求小，但同时同化结果更好，且耗时更短，条件允许的情况下宜采用同时同化的方法。

如前所述，电离层驱动参数对电离层预报是至关重要的。仅同化获取电子密度，没有优化的电离层驱动参数，难以对电离层进行可靠的预报。本文构建的电离层同化模型，只可用于电离层现报，接下来我们将进行电离层驱动参数的同化预报研究。此外，电离层同化预报还涉及到观测数据的质量控制、超大矩阵的存储与计算、误差协方差的选择等问题，还需对算法进行进一步优化，以改善同化结果的精度、提升同化速度。

致谢

感谢 COSMIC Data Analysis and Archive Center 提供 COSMIC 电离层掩星数据，感谢 NCAR 提供 TIEGCM 模式。

参考文献

[1] Scherliess, L., D.C. Thompson, and R.W. Schunk, Ionospheric dynamics and drivers obtained from a physics‐based data assimilation model. Radio Science, 2009. **44**(1).

[2] Yue Xinan, Modeling and Data assimilation of Mid- and Low-Latitude Ionosphere. [D]. 2008, wuhan Institute of Physics and Mathematics, Graduate University of Chinese Academy of Sciences)（乐新安，中低纬电离层模拟与数据同化研究 [D]. 2008，中国科学院研究生院，武汉物理与数学研究所）.

[3] Schunk, R.W., et al., Global assimilation of ionospheric measurements (GAIM). Radio Science, 2004. **39**(1).

[4] Scherliess, L., et al., Development of a physics‐based reduced state Kalman filter for the ionosphere. Radio Science, 2004. **39**(1).

[5] Scherliess, L., et al., Utah State University Global Assimilation of Ionospheric Measurements Gauss‐Markov Kalman filter model of the ionosphere: Model description and validation. Journal of Geophysical Research: Space Physics (1978–2012), 2006. **111**(A11).

[6] Matsuo, T., I.T. Lee, and J.L. Anderson, Thermospheric mass density specification using an ensemble Kalman filter. Journal of Geophysical Research: Space Physics, 2013. **118**(3): p. 1339-1350.

[7] Lee, I., et al., Assimilation of FORMOSAT‐3/COSMIC electron density profiles into a coupled thermosphere/ionosphere model using ensemble Kalman filtering. Journal of Geophysical Research: Space Physics (1978–2012), 2012. **117**(A10).

[8] Evensen, G., The ensemble Kalman filter for combined state and parameter estimation. Control Systems, *IEEE*, 2009. **29**(3): p. 83-104.

[9] Yue, X., et al., Error analysis of Abel retrieved electron density profiles from radio occultation measurements. Annales Geophysicae, 2010. **28**(1): p. 217-222.

[10] Bust, G., T. Garner, and T. Gaussiran, Ionospheric Data Assimilation Three‐Dimensional (IDA3D): A global, multisensor, electron density specification algorithm. Journal of Geophysical Research: Space Physics (1978–2012), 2004. **109**(A11).

[11] Yue, X., et al., Data assimilation of incoherent scatter radar observation into a one-dimensional midlatitude ionospheric model by applying ensemble Kalman filter. Radio Science, 2007. **42**(6): p. n/a-n/a.

[12] Yue X A, Wan W X, Liu L B, et al. Development of an ionspheric numerical assimilation nowcast and forecast system based on Gauss-Markov Kalman filter—An observation system simulation experiment taking example for China and its surrounding area. Chinese J. Geophys. (in Chinese), 2010，53(4):787~795（乐新安，万卫星，刘立波等. 基于 Gauss - Markov 卡尔曼滤波的电离层数值同化现报预报系统的构建——以中国及周边地区为例的观测系统模拟试验. 地球物理学报, 2010，53(4):787~795）.

[13] Hmaill, T.M., Ensemble-Based Data Assimilation. Workshop on Predietability ECMWF, 2002.